WHAT IS CLAIMED IS:

1. A computer-readable medium having stored thereon a first lexicon data structure for each of a plurality of lexicon words, the first lexicon data structure comprising:

a host form variant field containing data representing a host form variant;

a host form word field containing data representing a host form word for the host form variant represented by the data of the host form variant field; and

a verification field containing data representing a property of the host form variant represented by the data of the host form variant field, the property being indicative of whether the host form variant is itself a valid word or whether the host form variant must be combined with another entry in the lexicon.

2. The computer-readable medium of claim 1, wherein the first lexicon data structure further comprises a segment association field containing data associating the host form variant with certain types of attachment entries in the lexicon to define valid combinations between the host form variant and at least one of the attachment entries in the lexicon.

3. The computer-readable medium of claim 2, wherein the plurality of lexicon words correspond to verb-clitic forms, the data contained in the host form word field of the first data structure represents a verbal host form, and the data contained in the host form variant field of the first data structure represents the host form variant of the verbal host form.

4. The computer-readable medium of claim 3, wherein the certain types of attachment entries in the lexicon are clitic attachment entries in the lexicon.

5. The computer-readable medium of claim 4, wherein the certain types of attachment entries in the lexicon are clitic pairs attachment entries in the lexicon in which each clitic pair is represented in the lexicon as a single unit.

6. The computer-readable medium of claim 5, wherein the plurality of lexicon words correspond to Spanish language verbs, and wherein the certain types of attachment entries in the lexicon are Spanish clitic types.

7. The computer-readable medium of claim 2, having further stored thereon a second lexicon data structure for each of the attachment entries in the lexicon, the second lexicon data structure comprising a segment association field containing data associating, as valid word forming combinations, a corresponding attachment entry in the lexicon with host form variants having matching data in their respective first lexicon data structure segment association fields.

8. The computer-readable medium of claim 7, wherein the second lexicon data structure further comprises a final segment field containing data indicating whether the corresponding attachment entry must appear in a final position of any words formed from a combination of a host form variant in the lexicon with one or more attachment entries in the lexicon.

9. The computer readable medium of claim 7, wherein for attachment entries in the lexicon representing two words, the second lexicon data structure further comprises a word break segmentation field containing data indicative of a word break location.

10. The computer-readable medium of claim 1, wherein if the host form variant and the host form word are identical, then the first lexicon data structure omits the host form word field.

11. A text analyzer including the computer readable medium of claim 1.

12. The apparatus of claim 11, wherein the text analyzer comprises a word breaking component.

13. The apparatus of claim 11, wherein the text analyzer comprises a spell checking component.

14. The apparatus of claim 11, wherein the text analyzer comprises a grammar checking component.

15. A method of annotating verb-clitic form segments in a lexicon, the method comprising:

defining, for a segment, final segment data indicative of whether the segment must appear in a final position of any verb-clitic words formed using the segment.

16. The method of claim 15, and further comprising:

defining, for the segment, segment association data indicative of valid combinations of the segment with other types of segments to form verb-clitic words.

17. The method of claim 16, wherein the segment is a clitic form segment, and wherein the step of defining segment association data further comprises defining the segment association data such that it is indicative of classes of clitic hosts variants to which the clitic form segment can appropriately be attached.

18. The method of claim 16, wherein the segment is a clitic pair form segment, the method further comprising:

defining, for the segment, word break data indicative of a word break location in the clitic pair form segment.

19. The method of claim 16, wherein the segment is a clitic host variant, the method further comprising:

defining, for the segment, verification data indicative of whether the clitic host variant must be combined with a clitic segment to form a valid word.

20. A method of combining first and second verb-clitic form segments from a lexicon to form a verb-clitic word, the method comprising:

determining whether absence of final segment data associated with the first verb-clitic form segment indicates that the first verb-clitic form segment cannot be a final segment of the verb-clitic word;

determining whether final segment data associated with the second verb-clitic form segment indicates that the second verb-clitic form segment must be the final segment of the verb-clitic word; and

combining the first and second verb-clitic form segments from the lexicon to form the verb-clitic word only if it is determined that the first verb-clitic form segment cannot be the final segment and

that the second verb-clitic form segment must be the final segment.

21. The method of claim 20, and further comprising:

determining from clitic class data associated with the respective first and second verb-clitic form segments whether the first and second verb-clitic form segments share at least one of a plurality of different clitic class associations; and

combining the first and second verb-clitic form segments from the lexicon to form the verb-clitic word only if it is determined that the first and second verb-clitic form segments share at least one of the plurality of different clitic class associations.